



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Q&A: Uses and challenges of 'big data' for agricultural development

Chair: Dr Kim Ritman

Australian Chief Plant Protection Officer and Chief Agricultural Scientist



Steve Mathews, Mario Herrero and Ken Street in the Q&A for session 2.

Q: Frances Hoyle, *The University of Western Australia*

One observation and one question. Observation: I'd like to draw a parallel between Dr Sibanda's presentation yesterday and her emphasis on her grandmother and gender, and 'big data' this morning where gender has been rather obviously absent. So what is the role of gender in 'big data' for smallholders, not just for the small guys, but also for the small women? What is the role of gender knowledge in genebanks? What's the role of women in Ghana in relation to drought insurance? And what do we know about users who have apps, both men and women, who try to find out about how to buy trades, market opportunities and weather events? I would be delighted to hear from you, thank you.

A: Steve Mathews

Honestly, it's something that I haven't thought about a lot. We market a product that is available to everyone who is interested in using it, so in terms of different treatments of the genders with regard to 'big data', it's a problem that we haven't really addressed, honestly.

A: Mario Herrero

Yes, the role of women in 'big data' – enormous, I think, and especially because if you look at smallholder systems the vast majority of the poorest of the

This report of the Q&A has been prepared from a transcript.

smallholders would actually be female-headed households. So just from that perspective, not in the project that I spoke about but in other projects, we've been creating something we are calling 'data rescue exercises of farming systems data'. This is allowing us to do much better analyses of the roles of women, and also much better analysis of how to target technologies so that when you move to the more formal markets they're not swamped by the men, for example, which is something that often happens in parts of Africa. And Dr Sibanda also mentioned something else that is really important: we know that female-headed households have better nutrition outcomes for kids and for the family because they tend to use the resources differently. So people working in the health sector are trying to create the link between women farmers and the diversity of what is produced and the health metrics. To me it seems obvious that when we can involve women enumerators and women doing the analysis, the perspective will also be slightly different to what we otherwise see. It also happens when there are women scientists in the field, and gladly I see here a lot of young females, probably very eager to go and start engaging with these people.

Q: John Muir, currently consulting for Oxfam in Cambodia

My work involves conflicting roles in Cambodia, as a consultant in Oxfam's resilience program for small landholders, and also consulting on large-scale concessional land. Alongside that I've come across articles that report Monsanto is about to buy ag data and then sell it to everyone – is that true or not?

A: Ken Street

I have not heard of that. Clearly, Monsanto's a big player, but it is just another one of the big players among many.

Q: Shumaila Arif, Charles Sturt University

My question is for Mario Herrero. Do you think that agricultural diversity is traded off with biosecurity? So when we are addressing smallholder farmers should we address biosecurity first and then go for mixed farming?

A: Mario Herrero

Well, I think there is less of a trade-off, because if you have a large range of species you can deal with risk and with biosecurity risks much more effectively than if you have only one or two crops, which is the more common situation. And if you are implementing integrated pest management in mixed systems, well that actually deals with biosecurity issues pretty well in many cases.

Q: Peter Wynn, Charles Sturt University

We've talked about patterns of food production, but one of the major limitations of the world is the distribution of food. How can we best map – and I address the question to Steve and Mario – how can we best map the distribution patterns of food throughout the world so that we can identify the bottlenecks that limit our ability to deliver food to those who really need it, because that, as I see it, is a major issue.

A: Steve Mathews

The land is trickier to model than production, there's no question about that. We're working hard on it, and there is data out there. It's not always easy to get; it's in a variety of formats involving point-of-sale systems, and I realise I'm talking about developed markets when I say that. But that sort of technology is penetrating down into the less-developed markets and becoming more available. And also, governments are collecting information on that stuff. But you have a good point: our grasp of distribution and demand is not as good as our grasp of supply and production.

A: Mario Herrero

Can I add to that? There are a couple of products of bilateral trade. They're not very spatially explicit, but at least you would have trade flows between countries of different commodities and different food products as well, even processed food and so on. So, for example, the University of Kassel will have one, and the University of Vienna would have another one that is very widely used by the integrated assessment community.

Q: Joseph Macharia, Queensland University of Technology

My question is to Mario. There is a trend of declining numbers of smallholders according to the latest studies, and there is now the emerging medium-holder, who holds around five acres. Do you think these farmers are the best suited to adapt to these technologies, given that there are also so many groups using social media to share knowledge and get information? Medium-scale farmers, where there are declining numbers of smallholder farmers, especially in some countries in Africa such as Zambia, Ghana, Malawi and others?

A: Mario Herrero

Yes, I think that is a really important target group and I think it's a target group that is probably emerging also for most of the donors. We are finding that as farm-size shrinks in these countries, there are not enough incentives among the really small farmers (of 2 hectares or so) to actually invest in the land, so they end up having to do a range of other things. Yes I would agree with you that the engine of growth in the smallholder sector would be farmers slightly bigger than the typical 2-hectare smallholder that you would see in East Africa and similar regions – bigger being holdings in the range of around 5–10 hectares.

Q: Sam Coggins, Agricultural Science student, The University of Sydney

G'day. I specialise in data analysis. Steve mentioned in his presentation about the limitations of data and that it's often a bit of a mess, it's not uniform, or it might not be accurate or precise, and that's obviously a limitation when you're trying to map farm-size or find useful germplasm. I was wondering what your key strategies are for overcoming those limitations and making unreliable data useful?

A: Steve Mathews

In our case, it's just a lot of hard work. We have a philosophy of not actually changing data that has come from a third source, or a second source, depending how you look at it. Instead we deal with the source to get it corrected there,

so that there aren't different versions of that data being propagated around the world. It takes time and it takes a lot of effort. In theory, that is what our business is about, to some extent.

A: Ken Street

In our case with FIGS, the really difficult thing is to get accurate latitudes and longitudes for the collection sites; there's a lot of fuzziness there. So what we would do is give a metric to what we think the accuracy is and take that into account when we're building our sets.

A: Mario Herrero

From the agricultural data that we use, I can tell you that there's a fair bit of science but there's also a fair bit of 'black arts' that we use in putting this together, in making all the necessary judgements of what is good data for certain parts and so on. I think there's a real missed opportunity, and I hope that Andy will talk about this later, in that we always have to go around chasing people, or sometimes we find data serendipitously. It would be really nice if we had this well consolidated in proper spaces, well maintained, etcetera. A lot of effort goes into the cleaning and the maintaining of the data, and if these products were already in certain repositories, it would be really useful actually.

A: Steve Mathews

One thing I'd like to add, is that I know of a lot of small firms that are doing their own cleaning and organising of data – and they're siloed. They keep the work secret from each other because they see it as a competitive advantage. What a waste of time! If there was a group like GODAN or Gro or similar that did this, where it's done once and it's reliable, that would be a huge help to a lot of people.

Q: Sophie Lamond, *The University of Melbourne*

That's a good segue to my question. I suppose the projects that you've all presented are reasonably open and they are talking to Commonwealth resources. But we need to be realistic – data is a tradeable commodity and there are a lot of people for whom their individual data can go into something open and good, or it can become a tradeable competitive advantage for profit. How are we setting about empowering smallholder farmers to understand issues about their personal advocacy, privacy and the power that they actually have with their own data, and how do we have these conversations?

A: Steve Mathews

I've been involved in a lot of discussions about how to get data from smallholders in a reliable way, and my answer, and other people's as well, has frequently been, "Why don't you pay them?". And people act as if that's a terrible idea and it's somehow morally wrong. I don't understand that. You're asking people for something of value and, in my opinion, they should get something of value in return. Frankly, the value could be quite small and yet still be appreciated and a reliable way of getting decent data. Clearly some people would provide poor data, but it would washout in the mass of data that you accumulated.

A: Ken Street

Can I make a comment – ICARDA do a lot of surveys for data and they stayed away from paying people for data, because people will construct data depending on what they think you want.

A: Steve Mathews

That's not an issue. But with the free model you're trying to get data from people who have better things to do. And you're talking about not just a survey, you're talking about collecting it every week for the next 20 years, or something like that. They're just not going to keep up with it, and you end up with abandoned projects.

A: Mario Herrero

I think there are two things about this. On one hand, yes, data are very valuable, but some of us, especially when working with public funds, feel we need to – and sometimes we are contractually obliged to – provide the data and to make it open source. For some groups you work with now, such as the CG System and many other donors including the Gates and others, that is now standard. But from my perspective, what I've found is that the sharing of the data, even from a non-profit perspective, is what raises even more money for groups like mine, because I'm seen as the good guy always giving data out. What happens? I'm always included in new projects, for example if there is a new grant, a new paper – and that will lead to other grants, etcetera, new initiatives. I think without a doubt that open data is a much better solution. When you are trying to protect it, you will find that three people are after your data ... but thousands of people will use the open data, and your protected data will become obsolete in seconds. By the way things are going at the moment, I don't think that we're in a position to be able to hold onto information forever, because its value is changing constantly. It's getting a lot cheaper to do that analysis that I presented now we've done another campaign with five times the data, the crowd-sourced data. Now we can go and repeat it and it was done in two months. Data becomes obsolete very quickly at the moment.

Q: Wendy Umberger, *Centre for Global Food & Resources, Adelaide*

I am an ag-economist and we are involved in collecting data at household level and also using time-series data. So I want to throw out a question to any of you who can answer it, relevant to the previous questions. I don't believe data does become obsolete when you're looking at household-level data, and particularly when we're trying to look more at the connection between agriculture and nutrition at the household level. Yet the data at the household level, the nutrition data, is often using big balance sheet data that's a mess because of trade issues. Big assumptions are being made on the basis of data that's really quite poor. Public sector funds, across the board – the USDA was mentioned this morning and we see it in Australia – they're really pulling back in terms of collecting even household-level data. This is a policy conference, and I'd really like to hear some comment on how we can get governments to invest again in that household data at the farm level, and to improve the quality. Because to do good time-series analysis we need good quality continuous data sets and those

are falling apart, even in places like Australia and the US. The same goes for the data sets we have for the countries that we're all working in. So I'd love some comment on that, from a policy standpoint.

A: Mario Herrero

I completely agree with you and perhaps I shouldn't have said that data become 'obsolete' for current analysis, because definitely panel data are essential. Probably the best example that we have now is the World Bank LSMS data (Living Standards Measurement Study) for smallholders. It's not perfect but we need to try to come up with a model that actually does what the LSMS group are doing for a broader range of countries. Yeah, it's very expensive, but to do this systematically ... I think it's the only way to be able to get much better analysis and much better ground-truthing of the kinds of models that we are actually trying to implement.

And also, with the LSMS data, we find it needs to be open to more disciplines, because you start looking at it and then somebody from the livestock science area notes that they didn't collect, say, two particular variables, and those might have been two variables that would have really enhanced the value of the information, and so on. So even organising what kind of data we're going to need, and perhaps framing something around the kind of household data that we need for that if we're serious about the SDGs ... still the LSMS probably doesn't have everything that we need. Ideally we'd be taking the big initiatives, taking the big goals to actually drive our data processes, instead of trying to retrofit just what we have, and for example trying to use it to monitor progress towards the SDGs. There would have to be some kind of international panel to organise this for sure.

A: Steve Mathews

I'd like to suggest an alternative to that idea, which is obviously a great idea. What's going on in the US right now is instructive: namely that not only are our existing programs being cut back – in other words data is not being collected in quite the way it used to be – but also that existing data sets are being obscured from view, which is unbelievable! I think there's a lesson to be taken from this, and that is that policy is subject to change, and any solution that relies on policy-makers to maintain a constant view on any subject at all is a very dangerous course of action.

I would like to suggest as an alternative – probably an unpopular one – that, if these things actually work commercially, and they continue to work commercially, then they will continue. And what I mean by that is, for instance, we know that our Gro Intelligence product is not viable at the smallholder level, but we are priced low enough that it is viable at just above the smallholder level. Then that individual is free to disseminate the information to everyone he or she deals with. That is a 'sustainable' model – to use a term that's popular for describing government-related and charity-related things – unlike a model that relies on political parties sticking with something that they said they were going to do. Policies change. For a model that relies on people's interest (as in the example of the data collection at the household level – and I'm not familiar with

LSMS but I certainly agree that it's very important), would it hurt to ask people to exchange their data for something of value? I don't know what that would be, but why not think of it that way, instead of just asking for it – even though they frequently will give it to you. And by the way, when you're asking for data and you pay for it, you get a certain type of data, as Ken pointed out; but when you ask and you don't pay for it you also get a certain type of data. There are people who won't give data away for free, and so you don't get their data. This is something that I think people should think about a little more than they seem to be doing.

Q: Richard Dickmann, Bayer Crop Science Australia

A question for Dr Street. We heard last night and this morning of the importance of diversity in diets. A lot of that around the world is provided by secondary crops. To what extent does the FIGS system provide a database that can help with the breeding of these secondary crops?

A: Ken Street

FIGS isn't really a set of databases; it's more an approach. So if we have the data associated with those secondary and relict crops, orphan crops, we can certainly apply to fix it, and that applies to wheats or people, and it certainly applies to weird and wonderful crops. So the idea would be, with these small collections, to try and assemble as much accession-level data as possible, particularly the latitudes and longitudes, and then we can apply a FIGS system to the breeding process.

Chair

Thank you to all speakers in this session.